Transformations of Count Data for Tests of Interaction in Factorial and Split-Plot Experiments

By: Mark E. Payton, Scott J. Richter, Kristopher L. Giles, and Tom A. Royer.

Payton, M. E., Richter, S. J., Giles, K., L., & Royer, T. A. (2006). On Transformations of Count Data for Tests of Interaction in Factorial and Split-Plot Experiments. *Journal of Economic Entomology*, 99(3): 1002-1006.

Abstract:

In applied entomological experiments, when the response is a count-type variable, certain transformation remedies such as the square root, logarithm (log), or rank transformation are often used to normalize data before analysis of variance. In this study, we examine the usefulness of these transformations by reanalyzing field-collected data from a split-plot experiment and by performing a more comprehensive simulation study of factorial and split-plot experiments. For field-collected data, significant interactions were dependent upon the type of transformation. For the simulation study, Poisson distributed errors were used for a 2 by 2 factorial arrangement, in both randomized complete block and split-plot settings. Various sizes of main effects were induced, and type I error rates and powers of the tests for interaction were examined for the raw response values, log-, square root-, and rank-transformed responses. The aligned rank transformation also was investigated because it has been shown to perform well in testing interactions in factorial arrangements. We found that for testing interactions, the untransformed response and the aligned rank response performed best (preserved nominal type I error rates), whereas the other transformations had inflated error rates when main effects were present. No evaluations of the tests for main effects or simple effects have been conducted. Potentially these transformations will still be necessary when performing these tests.

Keywords: data transformation| experimental analyses | factorial arrangements

Article:

Recording observed responses as count variables (number of insects) in designed experiments is representative of most studies in applied entomological research. It is commonly held that problems arise in trying to use standard analysis of variance (ANOVA) techniques without doing something to account for the nature of the count response (Steele et al. (1997), p. 245; Freund and Wilson (2003), p. 240). The need to account for typical count response variability can be seen when the experimental design is simple. For example, suppose we have an experiment that is in a completely randomized design with ***t*** treatments and ***r*** replicates for each treatment, the response is a count variable, and we assume that this count can be modeled with a Poisson distribution. A trait of the Poisson distribution is the equality of the mean and the variance; thus, when a treatment effect is present, the variance for the treatments that yield large means will be greater than those with smaller means. This violates the assumption of equal variances and is commonly referred to as heterogeneous variances. The result of such a condition leads to the pooling of unequal variances into the mean square error in the ANOVA and potentially biased means comparison tests.

A popular remedy for the problems presented by a count response in an ANOVA setting is the use of a transformation. Many popular textbooks, including Steele et al. (1997), Ott and Longnecker (2001), and Freund and Wilson (2003), advocate the use of trans- formations as a remedy for violations of the ANOVA assumptions. A landmark paper by Box (1988) dis- cusses the general situation in which transformations are needed. He states that often, in the absence of normality, the treatment means and variances become functionally related, leading to violations of ANOVA assumptions. He suggests a general form of the trans- formation that in turn provides what seems to be independence for these two statistics. Two recommended transformations used for the count data scenario are the square root and the logarithm (log) transformations (Tukey 1977).

Although transformations tend to work well in some situations, difficulties arise when they are applied to factorial experiments or to experiments in which the treatment structures consist of a combination of factors [Fligner (1981), Blair et al. (1987), Thompson and Amman (1990)]. Examples of such experiments are split-blocks and split-plots, in which repeated measures experiments are a special case. Typically of interest in these experiments is the interaction among the various factors involved. The transformation that has garnered much attention regarding factorial experiments is the rank transform, i.e., replacing the data by their respective rank. For many years, the rank function has been used extensively in simple nonparametric methods, but it received special attention in papers by Conover and Iman (1976, 1981). They pro- posed that any parametric statistical procedure, such as ANOVA, can be changed into its nonparametric equivalent by applying the rank transformation to the original data or ranking the original responses. They advocated this approach for any experimental design. This certainly appealed to practitioners of statistics that were concerned with the assumptions associated with a parametric procedure. The rank transformation was easy to apply, and it did not require any additional computer software.

However, work by Fligner (1981), Blair et al. (1987), and Thompson and Amman (1990) showed that the rank transformation tends to induce interactions when none exist. Because it is a nonlinear trans- formation, the rank transform does not necessarily preserve the relationships observed in the original nontransformed data. In particular, it can be shown that in a factorial setting, the expected value of the rank of any observation is a function of the cell means (Blair et al. 1987). Thus, there can be interaction in the rank-transformed data when there is none in the raw data, especially when main effects are present in the raw data. The aligned rank transformation was offered as a solution to this problem. It attempts to remove the main effects, when present, from the calculated contrast used to assess the interaction rank test. For ex- ample, when testing for interaction ina2 by2 factorial arrangement, the “aligned” observations are *AYijk Yijk - MEAN (Ai) - MEAN (Bk),* where *MEAN* (*Ai*) is the estimated mean of level *i* of factor *A,* and *MEAN* (*Bj*) is the estimated mean of level *j* of factor *B.* Higgins and Tashtoush (1994), Richter and Payton (1999), and Mansouri et al. (2004) provide more details on the aligned rank transform for factorial arrangements in completely randomized and split-plot designs.

If the rank transformation induces interaction in factorial settings, can the same result occur with other transformations? If the experimental design includes multiple factors and a count response, does the square-root transformation perform well or does it suffer the same fate as the rank transform? Furthermore, if it does suffer from inflated error rates, would the aligned rank be a proper solution? Does the violation of ANOVA assumptions typically associated with count data make ANOVA tests for interaction unreliable? These potential problems are concerns for applied entomology studies that use factorial designs. In a recent survey of articles published in the *Journal of Economic Entomology* spanning a period of 6 mo, >13% involved transformed count data from studies evaluating a factorial arrangement of treatments (K.L.G., personal observation). Although this survey encompassed only one journal over a short time, it is evidence that entomologists frequently transform insect count data from factorial experiments.

In this study, we address inducement of interactions in factorial settings after transformation of count data by 1) providing the results of a reanalysis of recently collected field count data and 2) evaluation of a computer simulation study that explores the performance of several transformations often used in applied research studies.

Materials and Methods

Reanalysis Example. We illustrate concerns on the inducement of interactions in factorial settings by first presenting a simple example in which a transformation of count data might be used. Royer et al. (2005) examined the effect of planting date (early September, mid-September, or late September) of winter wheat in Oklahoma and the dosage of an insecticide applied as a seed treatment (imidacloprid at 0.00, 0.75, 1.50, or 3.00 lb/cwt) on aphid pest greenbug, *Schizaphis graminum* (Rondani), and the bird cherry-oat aphid, *Rhopalosiphum padi* (L.) counts. This test was con- ducted at two locations over 2 yr. However, for the purpose of our reevaluation, only the data from a late March sampling date at one location (Perkins, OK) is considered. During each year, four replicates were conducted, and the combination of replicate and year served as the blocking variable. The experiment was laid out as a split-plot arrangement in a randomized complete block design with planting date as the main unit factor and insecticide dosage as the split unit factor. By using PROC MIXED (PC SAS version 8.2, SAS Institute, Cary, NC), ANOVAs were conducted on the untransformed count data, and data were trans- formed by log(x + 1), square root, rank, and aligned rank procedures. We looked at the resulting probability values for testing the interaction of planting date and insecticide dosage to evaluate the relative value of each transformation.

Simulations. All simulations were performed using PC SAS version 8.2 (SAS Institute). Two different 2 by 2 factorial arrangements were used in the simulation, both of which were arranged in a randomized complete block design: a "regular" 2 by 2 factorial and a split-plot. Each situation has six blocks, called REPS in the simulations. For the 2 by 2 factorial arrangement, the following SAS code was used to generate the Poisson distributed response variables: Y RAN POI(0, REP + FACTORA + FACTORB + FACTORAB);. For the split-plot, the code was Y RAN- POI(0, REP + FACTORA + REP*FACTORA + FACTORB + FACTORAB);.

Note that the split-plot uses a replicate by factor A interaction, which serves to model the main plot error in a split-plot. To investigate the type I error rates (i.e., rejecting a true null hypothesis) associated with the tests of A by B interaction, various combinations of the effects of A and B were modeled in both the factorial and split-plot cases. Both of these factors had levels ranging from 0 to 10 modeled in the simulation. The FACTORAB term is defined to be zero. Five different responses were investigated: the untransformed values, the square-root transformation, the natural log transformation of the response plus 1, the rank transformation, and the aligned rank. Because of the easy export of parameters and the ability to suppress printing, PROC GLM was used for all analyses of variance conducted in the simulation study. For each model used, 10,000 iterations were generated for each combination of factors A and B.

For the transformations that performed well in the type I error simulation study, a power study was then conducted to examine the ability of the transformations to detect interactions present in the model. For this study, the main effects are defined as zero and the interaction (FACTORAB) had levels 1Ð10 modeled in the simulation.

Table 1. Probability values for testing the interaction of planting date and insecticide level by using transformed aphid count data from Royer et al. (2005)

Transformation	*P*
Aligned rank	0.0452
Logarithm	0.3790
Rank	0.5114
Square root	0.0817
None	0.0090

Split-plot ANOVA by using PROC MIXED, SAS Institute.

Results and Discussion

Reanalysis Example. Table 1 presents the *P* values for testing the interaction of planting date and level of imidacloprid for the transformed and untransformed aphid counts. The reported *P* values are all representative of the same original data set; the only difference was the transformation applied to the data before analyses. The decision to use a transformation, or choose which transform to use, makes a huge difference aphid counts. The reported *P* values are all representative of the same original data set; the only difference was the transformation applied to the data before analyses. The decision to use a transformation, or choose which transform to use, makes a huge difference in the conclusions that are likely to drawn from the analyses. For our data set, using the raw data or the aligned rank transformation would lead a researcher to conclude that a significant date by treatment interaction exists, but the decision to transform the data with either the log or rank would lead, in this case, to the opposite conclusion.

Use of the square root function as a transformation would put the researcher in a gray area regarding the presence of an interaction. With such contradictory conclusions, it is important to know which approach is appropriate. Is it necessary to transform in this situation? We address this question with the use of simulations.

Simulations. In Table 2, the type I error rates for testing the two-factor interaction for the factorial set- ting are summarized. In this situation, factors A and B are interchangeable, so the table does not contain all possible combinations of these factors due to symmetry. In Table 3, the type I error rates for testing the two-factor interaction for the split plot situation are summarized. Because factors A and B are not interchangeable in this case, all combinations are presented.

Table 2. Simulation type I error rates for testing the two-factor interaction in a factorial experimental design over increasing main effects for A and B

Magnitude of effect of factor A	Transformation	Magnitude of effect of factor B					
		0	2	4	6	8	10
0	Aligned rank	0.0527	0.0499	0.0548	0.0578	0.0570	0.0578
	Natural log	0.0495	0.0464	0.0341	0.0305	0.0314	0.0271
	Rank	0.0482	0.0525	0.0505	0.0481	0.0480	0.0464
	Square root	0.0504	0.0461	0.0396	0.0391	0.0390	0.0362
	None	0.0467	0.0485	0.0503	0.0536	0.0514	0.0485
2	Aligned rank		0.0532	0.0532	0.0593	0.0572	0.0592
	Natural log		0.0617	0.0741	0.0946	0.1084	0.1076
	Rank		0.0521	0.0480	0.0512	0.0501	0.0485
	Square root		0.0554	0.0595	0.0726	0.0735	0.0722
	None		0.0481	0.0484	0.0566	0.0532	0.0493
4	Aligned rank			0.0554	0.0565	0.0533	0.0588
	Natural log			0.1318	0.1985	0.2377	0.2824
	Rank			0.0392	0.0352	0.0370	0.0440
	Square root			0.0838	0.1114	0.1262	0.1380
	None			0.0503	0.0510	0.0502	0.0535
6	Aligned rank				0.0604	0.0583	0.0596
	Natural log				0.2899	0.3808	0.4590
	Rank				0.0251	0.0179	0.0213
	Square root				0.1459	0.1783	0.2120
	None				0.0534	0.0519	0.0530
8	Aligned rank					0.0609	0.0639
	Natural log					0.5153	0.6030
	Rank					0.0129	0.0090
	Square root					0.2325	0.2750
	None					0.0521	0.0569
10	Aligned rank						0.0613
	Natural log						0.7011
	Rank						0.0046
	Square root						0.3218
	None						0.0562

Table 3. Simulation type I error rates for testing the two-factor interaction in a split plot design

Magnitude of effect of factor A	Transformation	Magnitude of effect of factor B					
		0	2	4	6	8	10
0	Aligned rank	0.0527	0.0499	0.0548	0.0578	0.0570	0.0578
	Natural log	0.0495	0.0464	0.0341	0.0305	0.0314	0.0271
	Rank	0.0482	0.0525	0.0505	0.0481	0.0480	0.0464
	Square root	0.0504	0.0461	0.0396	0.0391	0.0390	0.0362
	None	0.0467	0.0485	0.0503	0.0536	0.0514	0.0485
2	Aligned rank	0.0587	0.0614	0.0508	0.0564	0.0583	0.0557
	Natural log	0.0516	0.1220	0.2257	0.3334	0.4267	0.4893
	Rank	0.0494	0.0853	0.1166	0.1201	0.0995	0.0700
	Square root	0.0499	0.0803	0.1149	0.1617	0.2046	0.2316
	None	0.0545	0.0577	0.0492	0.0564	0.0541	0.0529
4	Aligned rank	0.0652	0.0634	0.0575	0.0591	0.0565	0.0549
	Natural log	0.0538	0.1876	0.3979	0.5956	0.7254	0.8220
	Rank	0.0496	0.1483	0.2793	0.3577	0.3186	0.2417
	Square root	0.0486	0.1019	0.1827	0.2748	0.3512	0.4310
	None	0.0583	0.0616	0.0536	0.0543	0.0552	0.0530
6	Aligned rank	0.0688	0.0629	0.0679	0.0634	0.0607	0.0565
	Natural log	0.0588	0.2179	0.4886	0.7114	0.8501	0.9236
	Rank	0.0540	0.1876	0.4207	0.5862	0.6432	0.5783
	Square root	0.0498	0.1098	0.2296	0.3360	0.4549	0.5520
	None	0.0586	0.0564	0.0629	0.0583	0.0592	0.0540
8	Aligned rank	0.0779	0.0720	0.0677	0.0646	0.0636	0.0630
	Natural log	0.0593	0.2354	0.5498	0.7800	0.8995	0.9571
	Rank	0.0571	0.2111	0.5002	0.7097	0.8154	0.8331
	Square root	0.0492	0.1137	0.2489	0.3893	0.5208	0.6318
	None	0.0670	0.0617	0.0600	0.0584	0.0590	0.0571
10	Aligned rank	0.0760	0.0724	0.0644	0.0658	0.0588	0.0690
	Natural log	0.0643	0.2664	0.5788	0.8195	0.9287	0.9731
	Rank	0.0597	0.2396	0.5372	0.7762	0.8839	0.9262
	Square root	0.0520	0.1259	0.2628	0.4312	0.5764	0.6906
	None	0.0630	0.0635	0.0583	0.0612	0.0538	0.0628

A, main unit factor; B, split unit factor.

For the factorial design, all the responses have acceptable type I error rates (=0.05) when one or both of the factors have small main effects (Table 2). How- ever, when both main effects are present, the log and square root transformations have highly inflated error rates. Under this condition, the rank transformation has the unusual characteristic of having a diminished error rate Simulations with the untransformed data and the aligned rank transformed data result in only slightly inflated error rates and would be deemed acceptable, especially compared with the alternatives. Results are very similar for the split-plot design (Table 3); however, evaluation of the rank transformation reveals excessively high error rates for interactions as main effects increase. Simulations for both the un- transformed data and aligned rank data from a split- plot design reveal only slightly higher than expected error rates, but both the square root and log transformations perform poorly.

For the power analysis (Table 4), nonzero interaction effects were included in the model and the probabilities of rejecting the null hypothesis of no AB interaction were calculated for factorial and split-plot situations. Based on type I error rate simulations, the aligned rank, rank, and untransformed response functions were evaluated. Because neither the square root nor the log transformation performed acceptably in the type I error rate simulations (as main effects in- creased), they were not included in this portion of analysis. The rank transform was included, although an argument could certainly be made to exclude it as well. The aligned rank most consistently has the highest probability of rejecting the null hypothesis of the three responses investigated, although in general power of the untransformed data is close to that of the aligned rank. The rank transformation consistently performs more poorly in comparison.

Table 4. Simulation results of probabilities of rejecting the null hypothesis of no AB interaction when interaction effects are included in the model for factorial and split plot experiments

Experimental design	Magnitude of AB	Transformation		
		Aligned rank	Rank	None
Factorial	1	0.0958	0.0940	0.0891
	2	0.2250	0.2084	0.2175
	3	0.4030	0.3593	0.3901
	4	0.5990	0.5217	0.5841
	5	0.7577	0.6723	0.7496
	6	0.8811	0.7898	0.8756
	7	0.9445	0.8666	0.9428
	8	0.9772	0.9161	0.9756
	9	0.9901	0.9436	0.9906
	10	0.9962	0.9630	0.9958
Split-plot	1	0.0941	0.0933	0.0899
	2	0.2211	0.2046	0.2094
	3	0.4096	0.3630	0.3971
	4	0.5965	0.5228	0.5795
	5	0.7739	0.6855	0.7626
	6	0.8733	0.7856	0.8675
	7	0.9466	0.8682	0.9420
	8	0.9744	0.9139	0.9735
	9	0.9919	0.9424	0.9897
	10	0.9978	0.9633	0.9973

Recommendations. The results of the simulation study rely upon the assumption of Poisson distributed data. For count data, this assumption certainly seems to be valid. For experimental designs with a factorial arrangement of treatments, when the assessment of interaction is of interest, results of our simulations suggest that some of the transformations that have been relied upon in the past may not be working as expected. Please note, however, we have not ad- dressed the subject of testing simple or main effects. The use of transformations may still be warranted and necessary for these tests. For this article, we focused on tests of interactions and limited our examination to two-factor experiments. We can speculate that a three- or four-factor interaction will suffer the same fate, but we have developed no simulation evidence to support that conjecture.

If presented with count data in a factorial or split- unit arrangement, we recommend testing the interaction with the use of ANOVA techniques on the untransformed data or with the use of an aligned rank. All indications suggest that either of these approaches are sound from an error rate and power standpoint. One should note that using the raw data to test the interaction is doing so in violation of ANOVA assumptions. Our simulations were built with the Poisson distribution, so in some cases both normality and homogeneity of variance assumptions were violated. The results of the simulations indicate that ANOVA, at least in the case of the test of interaction, is very robust against departures from these assumptions. When subsequently testing main effects or simple effects, the researcher may use transformations as a remedy for the violation of ANOVA assumptions.

Acknowledgements

We thank R. Berberet and N. Elliott for critical review of this manuscript, and the referees and editors for valuable suggestions. This work was approved for publication by the Director of the Oklahoma Agricultural Experiment Station and supported in part under project OKL02334.

References Cited

Blair, R. C., S. S. Sawilowsky, and J. J. Higgins. 1987. Limitations of the rank transform statistic in tests of interaction. Commun. Stat. B16: 1133Ð1145.

Box, G.E.P. 1988. Signal-to-noise ratios, performance criteria, and transformations. Technometrics 30: 1Ð 40.

Conover, W. J., and R. L. Iman. 1976. On some alternative procedures using ranks for the analysis of experimental designs. Commun. Stat. A5: 1349 Ð1368.

Conover, W. J., and R. L. Iman. 1981. Rank transformations as a bridge between parametric and nonparametric statistics. Am. Stat. 35: 124 Ð129.

Fligner, M. A. 1981. Comment. Am. Stat. 35: 131Ð132.

Freund, R. J., and W. J. Wilson. 2003. Statistical methods, 2nd ed. Academic, San Diego, CA.

Higgins, J. J., and S. Tashtoush. 1994. An aligned rank transform test for interaction. Nonlin. World 1: 201Ð211.

Mansouri, H., R. L. Paige, and J. G. Surles. 2004. Aligned rank transform techniques for analysis of variance and multiple comparisons. Commun. Stat. 33: 2217Ð2232.

Ott, R. L., and M. Longnecker. 2001. An introduction to statistical methods and data analysis, 5th ed. Duxbury, Pacific Grove, CA.

Richter, S. J., and M. E. Payton. 1999. Nearly exact tests in factorial experiments using the aligned rank transform. J. Appl. Stat. 26: 203Ð217.

Royer, T. A., K. L. Giles, T. Nyamanzi, R. M. Hunger, E. G. Krenzer, N. C. Elliot, S. D. Kindler, and M. E. Payton. 2005. Economic evaluation of the effects of planting date and application rate of imidacloprid for management of cereal aphids and barley yellow dwarf in winter wheat. J. Econ. Entomol. 98: 95Ð102.

Steele, R.G.D., J. H. Torrie, and D. A. Dickey. 1997. Principles and procedures of statistics: a biometrical approach, 3rd ed. McGraw-Hill, New York.

Thompson, G. L., and L. P. Amman. 1990. Efficiencies of interblock rank statistics for repeated measures designs. J. Am. Stat. Assoc. 85: 519 Ð528.

Tukey, J. W. 1977. Exploratory data analysis. Addison-Wesley, Reading, MA.